

1 **ABSTRACT**

2 Data is captured from a web site or other data source. Data is extracted
3 from the web page using a data harvesting script or other data acquisition routine.
4 The extracted data is then normalized and stored in a database. If data cannot be
5 extracted from the web page, a copy of the captured web page is stored without
6 personal information contained in the web page. The data harvesting script is then
7 edited based on an analysis of the captured web page.